

Telugu Pronominal Anaphora Resolution

D. Chandra Mohan ¹ and M. Sadanandam (Supervisor) ²

¹ Department of CSE, Gitam University, Visakhapatnam, A.P. India.

² Department of CSE, Kakatiya University, Warangal –A.P. 506 009.

¹chanduthecm@gmail.com, ²sadanb4u@yahoo.co.in.

ABSTRACT

The concept of Anaphora Resolution arises from the use of the anaphors. The process of determining the antecedent of an anaphor is called anaphora resolution. Anaphora is a compound word comprising of *Ana* back or upstream and *Phora* meaning act of carrying. Anaphors and antecedents are said to be co-referential if they have the same referent in the real world. Most of the recent work in Anaphora Resolution was related to Hindi, Malayalam and Tamil. We have attempted to build a Rule Based System for Anaphora Resolution for the Telugu language. The system designed is mostly based on syntactic information with only certain semantic and morphological features. We make some syntactic cues for each Telugu pronoun (personal, Demonstrative, Indefinite, Interrogative, Reflexive etc.,) and based on these syntactic cues we make rules for the pronominal resolution. The system was evaluated on a limited set of data. The system has been tested for only pronominal Anaphora Resolution. The results depend mainly on the gender agreement. Including the gender information; the system could generate more accuracy, 58.19%, 57.3%, 80.5% and 48.14% for Personal Pronouns, demonstrative pronouns, Interrogative Pronouns and Reflexive pronouns respectively. The base system (with out gender agreement) gave an average of 48% accuracy on different pronouns.

Key words— Anaphora, Shallow Parser, Pronouns

1. INTRODUCTION

Anaphora resolution is a complicated problem in Natural Language Processing and has attracted the attention of many researchers. The approaches that have been developed, traditional (from purely syntactic ones to highly semantic and pragmatic ones), alternative (statistic, uncertainty-reasoning etc.) or knowledge-poor, offer only approximate solutions. Anaphora resolution takes place in the wider context of natural language processing (NLP), an enterprise that started in the early fifties. Research in algorithmic approaches to anaphora resolution started in real earnest in the seventies. Natural language generation systems, like natural language processing systems, must have an anaphora generation components. For proper translations in

Machine Translation, an interpretation of anaphors is important.

The concept of Anaphora Resolution arises from the use of the anaphors. Example1: The Empress did not like her dress. Here pronoun "*her*" is the anaphor and "*The Empress*" is the antecedent.

Anaphora, in discourse, is a device for making an abbreviated reference (containing fewer bits of disambiguating information, rather than being lexically or phonetically shorter) to some entity (or entities) in the expectation that the receiver of the discourse will be able to disabbreviate the reference and, thereby, determine the identity of the entity. Most importantly, anaphoric resolutions are used in information retrieval, Automatic Abstraction,

Machine Translation and Natural Language Interfaces.

Types of Anaphora

Depending on how the various sentences or different clauses in a sentence are interlinked, we may have different types of anaphora.

Pronominal anaphora:

The most widespread type of anaphora is the pronominal anaphora which is realized by anaphoric pronouns [4].

For example the pronouns are: Personal pronouns, Possessive pronouns, Reflexive pronouns, Demonstrative pronouns, Relative pronouns. All pronouns need not be anaphoric in nature.

Definite noun phrase anaphora:

Typical cases of definite noun phrase anaphora is when the antecedent is referred by a definite noun phrase representing either same concept (repetition) or semantically close concepts (e.g. synonyms, superordinates).

Example:

Computational Linguists from many different countries attended the tutorial.

One-anaphora:

one-anaphora is the case when the anaphoric expression is realized by a "one" noun phrase.

Example: If you cannot attend a tutorial in the morning, you can go for an afternoon one.

Types of anaphora according to the locations of the anaphor and the antecedent:

Intrasentential: anaphor and its antecedent are located in the same sentence.

For Example: Rohith asked harini how her mother was. (She refers to harini mother)

Intersentential: antecedent is in a different (preceding) sentence from the anaphor

2. ANAPHORA RESOLUTION IN TELUGU LANGUAGE

Most of the recent work in Anaphora Resolution was related to Hindi, Malayalam and Tamil. We have attempted to build a Rule Based System for Anaphora Resolution for the Telugu language. We make some syntactic cues for each Telugu pronoun (personal, Demonstrative, Indefinite, Interrogative, Reflexive etc.,) and based on these syntactic cues we make rules for the pronominal resolution.

Syntactic Cues of different Types of Anaphora:

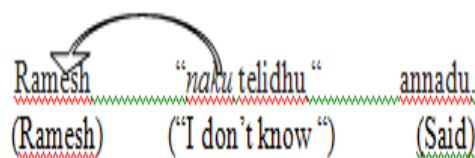
Pronominal Resolution

In anaphora resolution, pronominal resolution took priority over resolution of the non-pronominal. As far as the latter is concerned, resolution can be achieved using syntactic information alone, whereas in the case of the pronominal this is not possible. At the syntactic level, resolution can mean assignment of more than one candidate antecedent to the pronominal and this ambiguity can be resolved with the help of world knowledge.

There are different forms of Pronouns which are stated as below:

Personal Pronouns (PP):

Few of the pronouns which belong to this category are నను (nenu) (I / me), etc: - These are used to point to the first person.

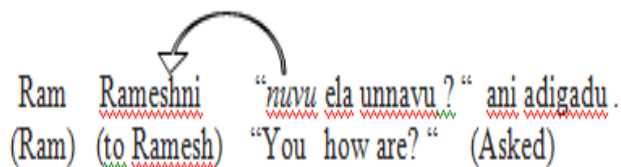


Example:

Syntactic cues: These pronouns may point to male and female.

Pronouns such as nenu, na usually are singular in nature, whereas pronouns such as memu can point to both singular (honorary notation) or plural.

□□□□□(nuvvu) (you) , □□□□(neevu) (you) ,
□□□□(meeru) (you) etc :- These pronouns are used
to point to the second person. Example:

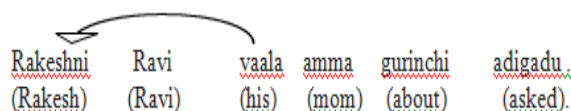


Syntactic cues: These pronouns may point to either
both male and female.

Pronouns such as nuvvu (you), nIvu (you) are
usually singular in nature, where as pronouns such
as mIru (you) can point be both singular (honorary
notation) and plural.

□□□□□(vaalu)(they) , □□□□(vaaru) (they) ,
(□□□)Adi (it): These pronouns usually point to the
third person, while vullu, vaaru points to Third
person in humans, Adi points to third person in
other living things.

Example:



Here the pronoun must point to the third person
which in this case is Rakesh.

Syntactic cues: These pronouns may point to both
male and female.

Few pronouns of this category may point to both
singular and plural, e.g., vaaru (honorary notation /
plural) while rest mostly point to plurals.

Pronouns such as "adi" may either point to human –
females or they may point to other living forms. In
certain cases they may not refer to any other proper
or common nouns.

Demonstrative pronouns (DP):

In English we have demonstrative pronouns which
stand alone, replacing rather than modifying the
noun.

For example, this is good.

In case of Demonstrative determiners, this usually
modifies a noun.

For example: I went into that house.

The most common examples of this kind are this,
that, these, those.

In Telugu, the pronouns which belong to this
category are mostly avi, ivi, adi (that/She/It), idi
(this/It) etc.

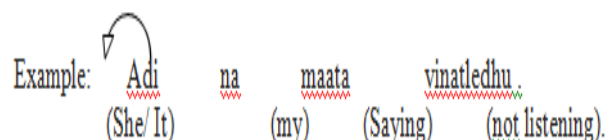
Similar to demonstrative pronouns, these may
replace the nouns.

For example: "Adi vinnu".

(That) (Listen to)

In English: Listen to that.

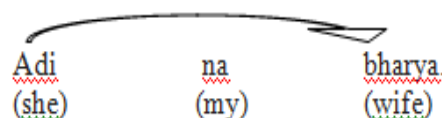
In this sentence Adi will not point to any other noun
(proper/common).



In case of Demonstrative determiners:

In English: She is not listening to what I am saying.

Here "adi" will point to either a noun
(proper/common) declared in the previous statement
or will point to NULL.



In English: She is my wife.

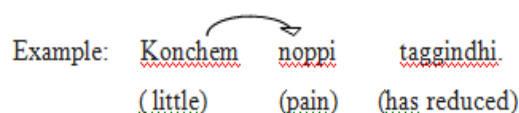
Here the pronoun will point to the common noun
which will follow it.

Syntactic Cues: They may point to living/non living
thing other than Human-males.

They generally point to the third person.

Indefinite Pronouns:

These pronouns refer to one or more unspecified
beings, objects, or places. Few of the pronouns which
belong to this



Categories are: Syntactic Cues:

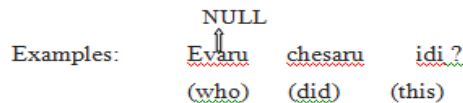
□□□□ □□□□ (Eavaro okaru) (someone) – It may
point to any human form (male / female), and it is
always singular.

□□□□ (Konchem) (some) – This usually is related to a measure of Quantity

□□□□□ (Konni)(some) - This usually points to Living / Non Living things, but is mostly never used for humans.

Interrogative Pronouns (INTRP):

An interrogative pronoun is a pronoun used in order to ask a question.

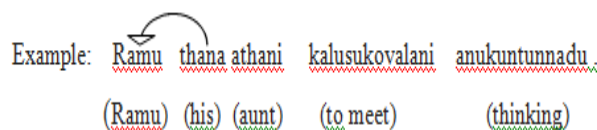
Examples: 

Syntactic Cues: These pronouns usually have the POS TAGS as WQ Usually used in a question.

Usually points to NULL.

Reflexive Pronouns:

These pronouns usually point to the nouns, adjectives or adverbs or pronouns which usually precede it. In Telugu, the closest to the English reflexive pronouns, such as “myself”, “yourself”, etc. is constructed using “thaanu” and “thaamu”.

Example: 

In English: Ramu is thinking to meet his aunt.

Syntactic Cues: These pronouns may point to male/female.

“thaamu” points to nouns which are plural while “tanu” points to nouns which are singular in number

Word order is very flexible in Telugu and these pronouns may point to nouns which are after the pronoun as well.

Relative and Correlative Pronouns:

A relative pronoun is a pronoun that introduces a relative clause. It is called a “relative” pronoun because it “relates” to the word that it modifies.

Inclusive Pronouns:

These pronouns are called inclusive as they include all. A few of the pronouns which belong to this category are “andaru”, “anni”.

Syntactic Cues:

These always point to nouns which are plural in number.

They do not require any gender agreement.

Certain pronouns e.g. “anni” will only point to Living or Non Living things, but never to Humans while some for e.g. “andaru” will always point to Humans.

Reciprocal Pronouns:

The reciprocal pronouns in English are *one another* and *each other*. Together with the reflexive pronouns *myself*, *yourself*, *ourselves*, *yourselves*, and others they are classified as anaphors. In case of Telugu, pronouns which belong to this type are “okariki okaru”, “mana madhyalo”.

Syntactic Cues: These pronouns will point to nouns with which they agree in gender and number.

System Design

The system designed is mostly based on syntactic information with only certain semantic and morphological features. When people speak natural language incorrectly, i.e., not strictly in accordance with rules of grammar and syntax, anyone can still make sense out of it. Hence developing an anaphora resolution rule set can always be incomplete. The most important features which we have considered in order to define the rules are

Number - Whether the pronoun will point to singular nouns or plural nouns.

Verb- Depending on the meaning of various verbs, certain rules have been crafted as shown below for anaphora resolution.

Inflexions- Shallow parser (explained below) helps to identify the inflexion which also helps in crafting the rules.

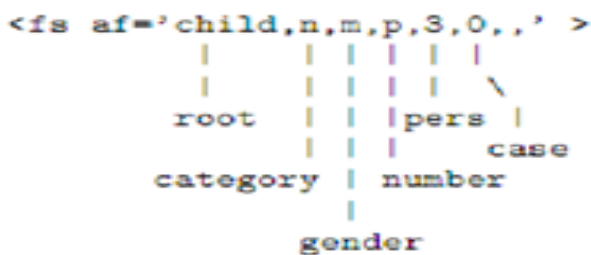
Gender Agreement- In Telugu, certain pronouns may point to both male and female while for certain pronouns, they usually point only to one kind.

We have considered having a Baseline system which does not include the Gender information, which was later improved upon by adding the gender information.

Telugu Shallow Parser:

The shallow parser gives the analysis of a sentence in terms of morphological analysis, POS tagging, Chunking, etc. Apart from the final output, intermediate output of individual modules is also available. All outputs are in Shakti Standard Format (SSF).

In the output, each line represents a word/token or a group. For each group, the symbol used is '('. Each word or group has 4 parts. The first part stores the tree address of each word or group, and is for human readability only. The word or group is in the second part, with part of speech tag or group/phrase category in the third part. Feature information is provided by the fourth part. Frequently occurring attributes (such as root, cat, gender, etc.) may be abbreviated using a special attribute called 'af' as follows:



Rule based Approach: Rules for Pronominal Resolution

Based on the Syntactic Cues of the various pronouns and the features extracted using Shallow Parser, the different rules are drafted as follows.

Personal Pronouns: In case of Personal Pronouns we have made different categories. These are as follows:

Sentences which do not have any verbs:

మహాత్మాగాంధీ చాలా గొప్ప వ్యక్తి. ఆయన తల్లి పేరు పుతలిబాయి.
(Mahatma Gandhi) (chala goppa) (Vyakthi) (Ayana) (thalli) (peru) (puthalibai).

(Mahatma) (Gandhi) (great) (person). (His)
(mother) (name) (Puthlibai).

For these sentences, 1st sentence has no pronoun, so we do not have to resolve any. Where as in the

Second sentence, absence of verb brings it under the category where the pronoun will point to the most recent noun added.

In case of system with gender agreement, our results would be better as in the following example. Consider the paragraph:

రాము ఒక మంచి బాలుడు.
(Ramu) (is a) (good) (boy).
అతని అక్క పేరు విజయ.
(his) (sister) (name) (vijaya).
ఆమె చాలా అమాయకురాలు.
(she) (is very) (innocent)
అతను ఆమె కన్నా చిన్నవాడు.
(he) (she) (to) (younger)

There are no verbs present in these lines. So they belong to the basic category. The pronouns will point to the most recent common/proper noun added. Here if we do not consider the gender then the pronoun in the 4th line (He points to Vijaya, which is incorrect. By considering the gender information, we may resolve this issue.

Improved solution:

రాము ఒక మంచి బాలుడు.
(Ramu) (is a) (good) (boy).
అతని అక్క పేరు విజయ.
(his) (sister) (name) (vijaya).
ఆమె చాలా అమాయకురాలు.
(she) (is very) (innocent)
అతను ఆమె కన్నా చిన్నవాడు.
(he) (she) (to) (younger)

For sentences which include non-conversational verbs: These verbs include for walking, listening, going etc. When the verbs of non-conversational type are present, then the gender of the verb is used to determine the gender of the subject.

For example: ఈ రోజు రాము వచ్చాడు.

(Today) (Ramu) (Came).

For example: ఇంట్లో సీత వచ్చింది.

(Yesterday) (Sita) (Came).

So here based on the feature information given by the shallow parser which is:

Here the gender of the verb is considered to be the gender of the proper noun.

For these, the rules applied are similar to the 1st category and would almost follow the same rules. Also, the tense of the verbs is to be considered.

Example: రమేశ్ మా ఇంటికి వచ్చాడు.
(Ramesh) (our/my) (home) (came).

Here the pronoun is for Personal Pronouns (1st person) where as the verb is of the third person. Hence the pronoun will not point to Ramesh.

Consider the following example: రమేశ్ అతని ఇంటికి చేరాడు.
(Ramesh) (his) (house) (reached).

In this example, the pronoun points to the 2nd or 3rd person, and since the verb is also in 3rd person, the pronoun will point to Ramesh.

Sentences which are within quotes:

Initially the pronouns outside the quotes are resolved, followed by the ones within the quotes. Before resolving the pronouns within the quotes, the pronouns outside the quotation marks are considered to determine the speaker or listener. All proper nouns declared within quotes point to second person.

For example: రాజీవ్ విజయ్ ని కలిసాడు.
(Rajeev) (Vijay) (met).
రాజీవ్ వాడిని ఎలా అన్నావు అడిగాడు.
(Rajeev) (him) (how) (was) (asked)
అతను "నేను బాగున్నాను" అని చెప్పాడు.
(He) (I) (fine) (that) (told).

For pronouns such as "□□□"," □□□□": they point to speaker or listener depending on whether their number and gender matches. In the following example we see that the structure of the sentences are the same but the same pronoun points to Listener in 1st example and to Speaker in the second, based on its

vaccAdu	VM	<fs af='vaccu,v,m,sg,3,,A,A' name="vaccAdu">
vacciMxi	VM	<fs af='vaccu,v,fn,sg,3,,A,A' name="vacciMxi">

gender and its number. Example1:

□□□□ □□□□ □□□□ □□□□.

(ramu) (wife) (name) (Anjali).

□□□□ □□ □□□□□□ □□□□ □□□□.

(she) (her) (brother) (likes).

Example2: అంజలి తమ్ముడు పేరు చండు. అమెకి వాడంటే చాలా ఇష్టం.
(Anjali) (brother) (name) (chandu). (she) (him) (very) (likes).

Demonstrative Pronouns: These pronouns usually include the types "□□□"," □□□"," □□□□□□"," □□□□□□":

For pronouns "□□□"," □□□": These pronouns may point to Living/Non living things but never to humans-males. They are singular in number.

In case the sentence does not have a verb: These pronouns point to the noun which follows it having no inflexions.

In case a verb is present and it is of the kind "□□□"," □□□□□": These pronouns are non anaphoric in nature.

In all the other forms the pronouns could be considered as pointing to the most recent Human(female) or Other Living form or Inanimate thing declared.

For Example (*no verb present*):

ఇది మా ఇంట్లో.
(this) (my) (home).

For Example (*verb present*):

□□□ □□□□□□ □□□ ఇక □□□□□□ □□□□.
(that) (after hearing) (she) (anymore) (not crying).
“□□□(that)” does not point to any noun .

For Example (*for third kind*):

□□□ □□ □□□ □□□□□□□□□.
(She) (my) (saying) (not listening)

In English: She is not listening to what I am saying.

In this example the pronoun will point to some human (female) or other living animals which has been declared most recently.

For Example: కొంచెము నేపేగా ఉంది
(little bit) (pain) (is there)

Interrogative Pronouns: For these pronouns e.g. □□□□ (who), □□□ (what), □□□□□ (what), □□□□□□ (to whom), they will have no reference.

Inclusive Pronouns: These include pronouns like “□□□□□ (all)”, “ □□□□□□ (we all)”. For pronouns like “□□□□□ (all)”, “ □□□□□□ (we all)” resolved by pointing them to plural noun (human) declared most recently or collectively to both the first person , second person and the third person simultaneously.

రాజు పిల్లల్ని చూసి “రేపు మనందరం సినిమాకి వెళ్దాము” అని అన్నాడు.
Raju (at children) (seeing) “(tomorrow) (we all) (to movie) (go)” (that) (told).

In this example we see that the pronoun points to the 1st person and the collective noun as well.

For pronouns such as “□□□□□ (all)” – These pronouns are similar to those Category 2 indefinite Pronouns.

Reflexive pronouns: These have syntactic cues similar to those of personal pronouns and can be handled similarly.

Reciprocal pronouns: These pronouns will always point to the subject depending on the gender and number. The information used to resolve these pronouns POS Tags, the gender and the number.

For Example: రామ్ మరియు రమేష్ ఒకరికొకరు నిజం చెప్పుకున్నారు.
Ram (and) Ramesh (each-other) (truth) (told).

3. EXPERIMENTAL RESULTS

The system was evaluated on a very limited set of data. The system has been tested for only Pronominal Anaphora Resolution. The base system (without gender agreement) gave an average of 48% accuracy on different pronouns. The results depend mainly on the gender agreement. Including the gender information, the system could generate more accuracy as shown in Table 6.1 (based on the pronouns identified by the Shallow parser in the data).

Pronoun type	Accuracy (%)
Personal Pronouns	58.0
Demonstrative Pronouns	57.0
Interrogative pronouns	80.0
Reflexive Pronouns	48.0
The total accuracy of the system	60.75

Table 6.1

4. CONCLUSION AND FUTURE WORK

Analyzing the results obtained by applying our rules, the following issues may be identified as part of future work.

Incorporating the NER Tool for Telugu:

Determination of the gender and named entity etc., have been done manually by creating a database. Incorporating the NER Tool would lead to better results.

Cataphora Resolution:

Rules for Cataphora Resolution must be included in order to improve the system.

Noun-Noun, Noun-Article Anaphora:

These two types need to be included for improving the scores of the system. These are the other two different types of Anaphora Resolution which we have to consider in future.

Sophisticated Rules:

Determination of more sophisticated rules (in case of Reflexive pronouns) is to be done in order to improve the system. The experimental results clearly reflect that the point of reference for the reflexive pronouns is less than 50% and better rules must be drafted to resolve them.

Comparing the results obtained for Telugu with results obtained for Hindi: The system would be used for Anaphora Resolution in Hindi, to see how well it can adapt with different languages.

REFERENCES

1. D.Chandra Mohan. "Blog Mining and Emotion Argumentation", Masters Degree Thesis under guidance of Prof. Sivaji Bandhyopadhyay, Jadavpur University, Kolkata, 2012.
2. Kamlesh Dutta , Nupur Prakash , Saroj Kaushik , "Resolving Pronominal Anaphora in Hindi Using Hobbs' Algorithm" , Vol. III, International Research Publication House, pp. 609-616.
3. Josef Steinberger , Massimo Poesio , Mijail A. Kabadjov ,Karel Jezek "Two Uses of Anaphora Resolution in Summarization", Reciprocal Constructions", Vol. I, pp.1-25.
4. <http://www.cs.ucdavis.edu/~vemuri/Grammar/9.%20pronouns-1.pdf>.
5. Robert Caldwell, Harrison "A comparative grammar of the Dravidian or South-Indian family of languages", London, 1856.
6. Barbara. C, Lust. "Lexical anaphors and pronouns in selected South Asian languages". Charles Philip Brown. "A grammar of the Telugu language", 1857.